SPATIAL PROPERTIES OF SPEECH MOVEMENTS

Vincent L. Gracco and Anders Löfqvist
Haskins Laboratories, 270 Crown Street, New Haven, CT USA

ABSTRACT

If there is one characteristic of speech that has plagued speech production theorists for years, it is variability. Acoustic correlates of a given phoneme and by inference vocal tract configurations exhibit variability arising from a number of sources. The present experiment was designed to examine the degree of spatial variation among a limited set of phonetic segments. Results suggest that variability in vocal tract positioning may be sound dependent reflecting different degrees of perception/production stability.

INTRODUCTION

An issue of theoretical importance in speech production is to determine the precision at which articulatory actions are being controlled. One characteristic of speech, however, that has plagued the development of a realistic understanding of control precision is variability. Variation may arise from many sources and it's understanding is crucial to the development of a realistic perspective on speech motor control. A general perspective can be obtained from consideration of the structure and function of the human nervous system as an information processing device. As pointed out by von Nuemann [1] the nervous system is an analog device that is ideally suited for reliable operation not precision. In this context it can be suggested that articulatory performance is good enough without incurring excessive "costs" [2] with the degree of precision inherently dependent on the listener's ability to extract meaning from the speech code. Alternatively, variation may be related to certain articulatory/acoustic relations such as those reflected in Stevens quantal theory [3,4] which implicitly assumes that certain consonants should exhibit more or less articulatory variability as a function of the proximity of so-called primary articulators to sensitive regions of the vocal tract in which small changes in articulation have large acoustic consequences.

Evaluation of spatial precision assumes that speech movements may in fact have spatial targets associated with them. The concept of spatial targets for speech was suggested by Lashley [5] in discussing space coordinate systems for controlling serial movements such as those for speech. In spite of the intuitiveness of speech motion planning in a spatial reference frame, the notion of spatial targets for speech have received little attention. One reason for the limited attention, again, appears to be related to the presence of variability in the observable signal in which variable vocal tract shapes yield acceptable acoustic signals [6] However, as noted above, this limitation may be related more to an unrealistic expectation regarding the precision of the articulatory target for speech. A recent perspective has been offered by Guenther [7] in which spatial targets for speech are viewed as regions rather than points (convex hulls) in orosensory space and conceptually is more attractive than target points. While the available data is limited it is difficult to imagine that speech is not planned to some extent in a spatial coordinate frame since inappropriately placed articulators will produce seriously compromised sounds. The purpose of the present experiment was to examine the spatial variation of a few of the phonetic segments of the language to determine how speech movement control varies as a function of phonetic identity.

PROCEDURE

The experimental group consisted of four subjects (two males, two females). Movements of the lips, jaw, and four points on the surface of the tongue were obtained using an electromagnetic transduction device [8]. The tongue receivers were placed approximately 1 cm behind the tongue tip and spaced approximately 1 cm apart. Data were hardware low pass filtered (200 Hz) and sampled at 625 Hz (12 bit resolution). Following the digitization, the voltages were digitally smoothed (25 points with a 3 dB point at 18 Hz) and the voltages were converted to positions in the midsagittal plane of the device. All data were rotated to the subjects' occlusal plane.

Subjects repeated a number of words embedded in the carrier phrase "Say _____ again." ten times. In order to examine the spatial variation associated with different phonetic segments, the two dimensional positions of the four tongue receivers were obtained at the time of zero (or minimum) speed associated with the target acoustic segment [9].

RESULTS

Shown in Figure 1 is the average tongue shape estimated from a cubic spline interpolation of the four average receiver locations for one subject for the three phones /s/, /r/, and /ae/ along with one standard deviation bars. For these comparisons the variability reflects the variation associated with repeating each word ten times. The words represented are /s/-"sack", /r/-"rack", and /ae/-"had". In considering the spatial variation there are two points of note: the degree of variation is quite different for the three segments and the different tongue regions display different degrees of variability.

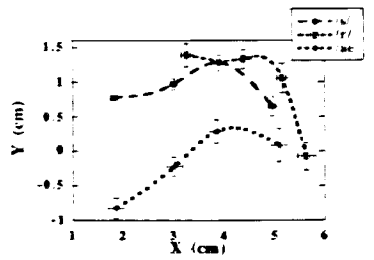

Figure 1. Derived tongue shape, from a cubic spline interpolation, during the steady-state or quasi steady-state behavior during the three phonetic segments /s/, /r/, and /ae/. The X dimension represents the subjects occlusal plane. Error bars reflect one standard deviation.

To examine the spatial variation in a more systematic manner the standard deviations in the spatial positions of the tongue receivers were obtained at the minimum speed associated with each phonetic segment of interest. The standard deviations in each spatial dimension were added providing an estimate of the average variation for each receiver location. For these comparisons the phone of interest was examined when it was in the initial position of the word. The words examined were: for /s/—"sack" and "sag"; for /r/—"rack" and "rag"; for /l/—"latter" and"ladder"; for /n/—"need" and "neat". The standard deviation in the X and Y dimensions were added for each of the 20 repetitions and are plotted as a function of the four phones reported here.

Figures 2 and 3 present the combined standard deviation in the X and Y dimensions for /s/ and /r/ (Figure 2) and /l/ and /n/ (Figure 3). As shown in Figure 1 there is a general trend for the variation at all positions to be greater for /r/ than for /s/ with a trend for the tongue front to show the smallest deviation compared to the tongue rear. Figure 3 shows the variation in tongue receiver positions for /l/ and /n/. The trend for these phones is for /l/ to show more spatial variation than /n/ for all receiver positions.

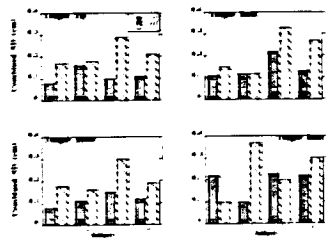

Figure 2. Combined standard deviations for /s/ and /r/ associated with each spatial dimension for each of the four tongue receivers for each of the four subjects.

While it might be concluded from these data that different phones differ in the degree of control precision required it should be noted that the results become more complicated when considering the same phones in different syllable positions. Shown in the next three figures are the estimated tongue shape for a single subject when the phones /l/, /r/, and /r/ were produced in different word/syllable positions. Figure 4 shows the tongue shape for /l/ in the words "ladder/latter" and "medal/metal". Figure 5 presents a similar comparison for /r/ when produced

word initial "need" and syllable medial "and". Interestingly, while /l/ demonstrated more spatial variation than /n/ in a similar context, the estimated tongue shape for /l/ is much more consistent across contexts than is /n/.

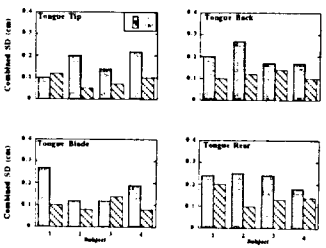

Figure 3. Combined standard deviations for /l/ and /n/ associated with each spatial dimension for each of the four tongue receivers for each of the four subjects.

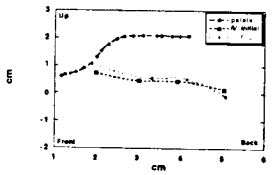

Figure 4. Estimated tongue shape for /l/ in two different syllable positions. Each receiver position reflects the average of 20 tokens (10 for each word). The top trace is an outline of the subjects hard palate.

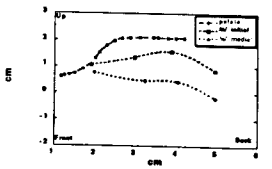

Figure 5. Estimated tongue shape for /n/ in two different syllable positions. Each receiver position reflects the average of 20 tokens (10 for each word). The top trace is an outline of the subjects hard palate.

DISCUSSION

Informed explanation of articulator variability rests on a number of assumptions regarding the control of speech movements; the degree of control precision and the goals for speech. At a conceptual level speech movements can be understood as goal-directed [10,11] and reflecting a level of control consist with obtaining changes in vocal tract configurations rather than movements of individual articulators [12,13]. The present results are consist with nervous system control operating on ensembles of articulators with differential degrees of precision depending on the context in which the variation is observed.

An example of the apparent looseness in the precision of articulatory control can also be found in recent simulation and synthesis results reported by Gay, Boe, & Perrier [14]. Parametric manipulation of vocal tract cross sectional area and constriction location was used to determine the acoustic and perceptual boundaries of certain isolated vowels. It was shown that the formants for each of the vowels were most sensitive to changes in cross sectional area compared to constriction location. Vowel perception, however, was insensitive to both manipulations. The results from Gay et al. [14] were somewhat at odds with the notion of the quantal characteristics of speech [3,4] suggesting rather that quantal regions may not necessarily be avoided because of the tolerance of the perceptual system. From these results it was concluded that the speech production mechanism has "...considerable latitude..." in specifying the articulatory targets. Limited kinematic data reported by Perkell and colleagues [15,16] is consistent with a relaxed degree of articulatory control.

In summary, the present report, while limited in scope, suggests that the specification of control precision can be thought of as an inherent property of each of the speech production units (phonetic segments) of the language. Moreover, the degree of variability may be systematically related to and ultimately reflect the perceptual tolerance of the language.

ACKNOWLEDGMENT

This work was supported by Grants DC-00121 and DC-00594 from the National Institute on Deafness and Other Communication Disorders, and in part by Esprit-BR Project 6975 - Speech Maps through Grant P55 from the Swedish National Board for Industrial and Technical Development.

REFERENCES

[1] von Nuemann, J. (1958). *The computer and the brain.* New haven: Yale University Press.

[2] Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, 46, 135-147.

[3] Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw-Hill.

[4] Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.

[5] Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffries (Ed.), *Cerebral mechanisms in behavior* (pp. 112-136). New York: John Wiley.

[6] Ladefoged, P., DeClark, J., Lindau, M., & Papcun, G. (1972). An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics*, 22, 48-75.

[7] Guenther, F. H. (1994). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Technical Report CAS CNS-94-012. Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems, Boston, MA.*

[8] Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992), "Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements", *J. Acoust. Soc. Am.*, vol. 92, pp. 3078-3096

[9] Löfqvist, A. & Gracco, V. (1994), "Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context", *Phonetica*, vol. 51, pp. 52-67.

[10] Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research*, 65, 156-166.

[11] Saltzman, E. L. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research*, Series 15, 129-144.

[12] Gracco, V. L. (1991). Sensorimotor mechanisms in speech motor control. In, H. Peters, W. Hultsijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering.* (pp. 58-78). North Holland: Elsevier.

[13] Gracco, V. & Löfqvist, A. (1994), "Speech motor coordination and control: Evidence from lip, jaw, and larynx interactions", *Journal of Neuroscience*, vol. 14, pp. 6585-6597.

[14] Gay, T., Boe, L.-J. & Perrier, P. (1992). Acoustical and perceptual effects of changes in vocal tract constrictions for vowels. *Journal of the Acoustical Society of America*, 92, 1301-1309.

[15] Perkell, J. S., & Nelson, W. L. (1985). Variability in production of the vowels /i/ and /a/. *Journal of the Acoustical Society of America*, 77, 1889-1895.

[16] Perkell, J. S., & Cohen, M. (1989). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *Journal of Phonetics*, 17, 123-133.